

METHOD AND SYSTEM FOR DATA EVALUATION, CORRESPONDING COMPUTER PROGRAM PRODUCT, AND CORRESPONDING COMPUTER-READABLE STORAGE MEDIUM

Patent number: DE10143712
Publication date: 2003-04-10
Inventor: REYMOND MARC A (DE)
Applicant: EUROPROTEOME AG (DE)
Classification:
- **international:** G06F19/00
- **european:** G06F17/30B, G06F19/00A
Application number: DE20011043712 20010830
Priority number(s): DE20011043712 20010830

Also published as:



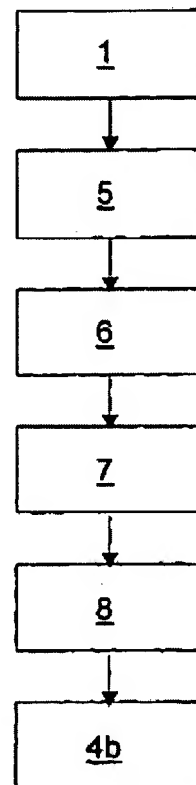
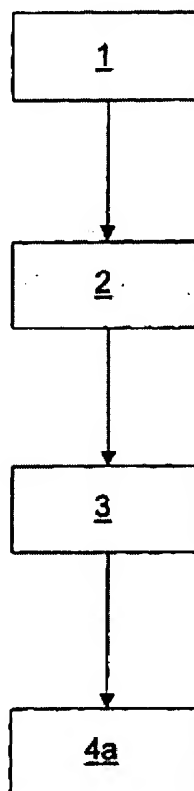
WO03021478 (A3)

WO03021478 (A2)

Abstract not available for DE10143712

Abstract of correspondent: **WO03021478**

The invention relates to a method and a system for data evaluation, to a corresponding computer program product, and to a corresponding computer-readable storage medium, which can be especially used as an internet-based, patient-specific prognosis system. In this case, clinical, pathological and molecular biological data can be integrated, and said data can be combined with relevant prognoses for a particular patient. The system thus enables an oncologist, for example, to decide on an individual treatment on the basis of a specific information pattern. The quality of the prognosis is improved by determining significant and secondary variables, leading to a clear reduction in the quantity of data to be evaluated, to the acceleration of the data evaluation and to the improvement of the prognosis.



Data supplied from the **esp@cenet** database - Worldwide

THIS PAGE BLANK (USPTO)

02P20510



①9 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENT- UND
MARKENAMT

⑫ **Offenlegungsschrift**
⑩ **DE 101 43 712 A 1**

⑤1 Int. Cl. 7:
G 06 F 19/00

②1 Aktenzeichen: 101 43 712.9
②2 Anmeldetag: 30. 8. 2001
④3 Offenlegungstag: 10. 4. 2003

DE 101 43 712 A 1

⑦1 Anmelder:
Europroteome AG, 16761 Hennigsdorf, DE

⑦4 Vertreter:
Anwaltskanzlei Gulde Hengelhaupt Ziebig &
Schneider, 10117 Berlin

⑦2 Erfinder:
Reymond, Marc A., 38350 Helmstedt, DE

⑤6 Entgegenhaltungen:
US 55 17 405 A

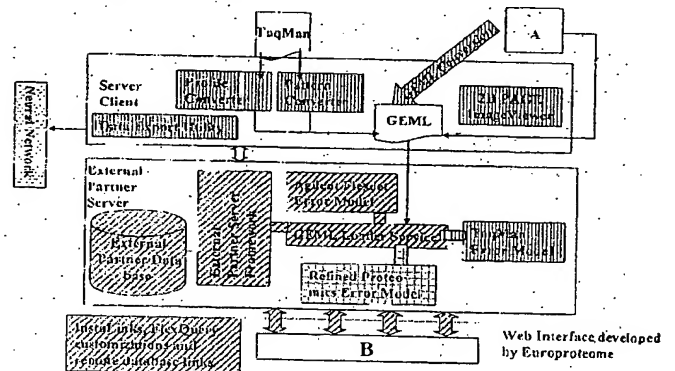
Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gem. § 44 PatG ist gestellt.

⑤4 Verfahren, Computersystem und Computerprogrammprodukt zur Datenauswertung

⑤7 Die Erfindung beschreibt Verfahren, Computersystem und Computerprogrammprodukt zur Datenauswertung, welche insbesondere einsetzbar sind als internetbasiertes patientenspezifisches Prognosesystem. Dabei wird die Integration von klinischen, pathologischen und molekularbiologischen Daten ermöglicht sowie die Verknüpfung dieser Daten mit relevanten Prognoseaussagen bei einem bestimmten Patienten. Dadurch erlaubt das System aufgrund spezifischer Informationsmuster, eine individuelle Therapieentscheidung zu treffen.

Die Erfindung wird u. a. die Schätzung des Metastatisierungsrisikos eines individuellen Patienten erlauben, so daß die Indikation einer adjuvanten Chemotherapie gezielter gestellt werden kann. Außerdem wird eine Schätzung der Probabilität der Therapieantwort bei verschiedenen Chemotherapeutika ermöglicht, so daß Tumoresistenzmuster erkannt werden können. Durch Selektion der molekularen Targets, die mit einem bestimmten klinischen Outcome verbunden sind, wird eine Reduzierung der Daten erreicht, die für eine Prognose ausgewertet werden müssen. Außerdem kann durch die Erfindung das Risiko von erfolglosen Therapieversuchen deutlich gesenkt, die Entwicklungskosten eines Medikaments reduziert und damit die Gesundheitskosten verringert werden.



DE 101 43 712 A 1

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren, ein Computersystem und ein Computerprogrammprodukt zur Datenauswertung, welche insbesondere einsetzbar sind als internetbasiertes patientenspezifisches Prognosesystem. Es wird in diesem Falle die Integration von klinischen, pathologischen und molekularbiologischen Daten ermöglicht sowie die Verknüpfung dieser Daten mit relevanten Prognoseaussagen bei einem bestimmten Patienten. Dadurch erlaubt das System beispielsweise einem Onkologen, aufgrund spezifischer Informationsmuster eine individuelle Therapieentscheidung zu treffen.

[0002] Die Informationstechnologie (IT) gewinnt in der Medizin zunehmend an Bedeutung. Dennoch wird derzeit die Patientenversorgung als Kernprozeß im Gesundheitswesen nur unzureichend unterstützt. Stattdessen stehen administrative Tätigkeiten im Vordergrund. Das Potential dieser Technologie erlaubt aber die Bereitstellung einer qualitativ besseren Patientenversorgung bei gleichzeitiger ökonomischer Nutzung der vorhandenen Ressourcen.

[0003] Einen wichtigen Bestandteil einer verbesserten Patientenversorgung bilden verlässliche Prognoseaussagen. Eine Prognose läßt sich jedoch nicht nur anhand allgemeiner Kenntnisse zu Krankheit und Patient erstellen; wichtig sind darüber hinaus auch Informationen zum vorherigen Verlauf der Krankheit bei jedem einzelnen Patienten. Hierbei spielen genaue klinisch-pathologische Daten und Informationen zur Nachsorge eine sehr wichtige Rolle. So muß hier beispielsweise die iterative Natur des Prognosebestimmung berücksichtigt werden.

[0004] Neben der genauen Kenntnis des Patienten und seiner Krankheit ist es ebenso wichtig, den vorliegenden Fall mit ähnlichen Fällen aus der Vergangenheit an eigenen oder externen Institutionen vergleichen und die damaligen Erfahrungen berücksichtigen zu können. Solche Vergleiche sind jedoch nur mit qualitativ hochwertigen Referenz-Datenbanken realisierbar.

[0005] Prognoseaussagen kommt insbesondere bei Krebskrankheiten eine große Bedeutung zu, um die bestmögliche Therapie zu bestimmen. Die Wichtigkeit ergibt sich aus der Tatsache, daß Krebs im Unterschied zu einer Viruserkrankung, welche in jedem Patienten die gleichen Symptome hervorruft, ein sich entwickelndes und individuelles Krankheitsbild hat. Bei diesen Krankheiten sind für Prognoseaussagen die folgenden Fakten von Bedeutung:

- Patienteninformationen wie Alter, Co-Morbidität, Zuverlässigkeit, usw.,
- Umgebungsinformationen wie Chirurg, initiale Behandlung, Versicherungssystem, Land, usw.,
- Tumorinformationen wie Pathologie, Tumorstaging, Mutationen, Genexpression auf Transkriptom- und Proteom-Ebene.

[0006] Um diese "Kunst der Vorhersage" zu erlernen, ist es wichtig vom Patienten auszugehen. Welche Fragen beschäftigen den Patienten? was will er vom Arzt wissen? Hierbei werden von Patienten - einer Umfrage zufolge - die folgenden Fragen am häufigsten gestellt:

- Inwieweit werde ich mit einer Behandlung geheilt?
- Wie ist die normale Lebensdauer mit einer Behandlung?
- Werde ich sterben, wenn ich nicht behandelt werde?
- Wie schnell wird sich die Krankheit ausbreiten, wenn ich nicht behandelt werde?

[0007] Die heute zu Verfügung stehende Staging-Systeme (wie z. B. das TumorNodeMetastasis-System der International Union Against Cancer - UICC) erlauben zwar bereits Aussagen für Patientengruppen, leider aber nicht für einen spezifischen Patienten. Bei der Prognose müssen aber Informationen auf jeden einzelnen Patienten bezogen werden unter Berücksichtigung von dessen spezifischer Situation, wohingegen bei der Diagnose das Spezielle verallgemeinert und übergangen wird. Künftig werden moderne Erkenntnisse aus der Tumor-Genexpressions-Forschung zusätzlich berücksichtigt werden müssen, um diesen Schritt von Diagnose zu patientenspezifischer Prognose - und damit eine individuelle Therapie - zu ermöglichen. Eine Aufgabe, die bisher noch nicht gelöst wurde.

[0008] Ein weiteres ungelöstes Problem bei den herkömmlichen Systemen besteht in der Bewältigung der erheblichen Datenmengen, die für eine qualitativ hochwertige Prognose ausgewertet werden müssen. Diese können vom Arzt (Onkologen) allein bei Therapie-Entscheidungen nicht mehr verwaltet werden. Auch die gegenwärtig verfügbare Rechentechnik und die für die Datenauswertung eingesetzten Programme sind nicht geeignet, um diese - insbesondere von den molekularbiologischen Datenbanken bereitgestellten - Datenmengen in vertretbarer Zeit auszuwerten.

[0009] Die Nachteile der vorhandenen Lösungen auf diesem Gebiet sind von führenden Krebsverbänden erkannt. Zum Beispiel werden alle Patienten mit kolorektalem Karzinom im UICC-Stadium 3 gemäss Richtlinien der Consensus-Konferenzen nach kurativer Chirurgie adjuvant behandelt, auch wenn nur 40% metachrone Fernmetastasen entwickeln. Dies führt zu überflüssigen Nebenwirkungen für die Patienten, und zu erheblichen Zusatzkosten für das Gesundheitssystem. Auf der anderen Seite werden 8% der Patienten im Stadium I und 14% der Patienten im Stadium II Fernmetastasen entwickeln, und bekommen laut Richtlinien keine adjuvante Chemotherapie, was zu einer erhöhten Krebslethalität führt (Köckerling, Reymond et al. J Clin Oncol, 1998).

[0010] Die Etablierung eines internet-basierten medizinischen Informationssystems für Ärzte ist als eine notwendige Unterstützung der ärztlichen Tätigkeit betont worden. Die American Cancer Society (http://www3.cancer.org/-cancerinfo/cancer_profiler.asp) und andere Organisationen wie die European School of Oncology ([www.cancerworld.org/progetti/cancerworld/start/pagine /Homeframe.html](http://www.cancerworld.org/progetti/cancerworld/start/pagine/Homeframe.html)), die University of Pennsylvania (www.oncolink.upenn.edu/resources/physicians) und andere (z. B. <http://www.cancer-home.com>) bieten schon jetzt Internet-Portale, um z. B. die Onkologen zu unterstützen.

[0011] Alle diese Portale geben jedoch nur die aktuellen Richtlinien und Empfehlungen der Experten-Konferenzen wieder und bringen deswegen keine Lösung zur oben erwähnten Problematik. Außerdem können diese Seiten von Patienten aufgerufen werden, was laut dem HON Code of Conduct (HONcode) für medizinische Webseiten im Gesundheitsbereich nicht geschehen darf: "Die Information auf der Website ist so angelegt, daß sie die existierende Arzt-Patienten-Beziehung unterstützt und keinesfalls ersetzt." (<http://www.hon.ch/HONcode/German/>).

[0012] Weiterhin gibt es bisher kein e-Health-Produkt auf dem Markt, das molekularbiologische Daten mit konventionellen klinischen und pathologischen Informationen verbindet, um den Arzt bei Therapieentscheidungen für Krebskrankheiten zu unterstützen. Man braucht nur die Vielfalt der möglichen Therapien zu erwähnen, und erhält die Menge an Informationen, die zu bewältigen sind, um die Notwendigkeit eines solchen Systems zu unterstreichen.

[0013] Ein Informations-System in der marktvorbereiten-

den Erprobungsphase, in dem für die Erzeugung der prognostischen Aussage neuronale Netze bzw. regelbasierte Systeme eingesetzt werden, ist bekannt. Hierbei werden einige tausend Datensätze (Sets) von prospektiven Patientendaten für die Auswertung berücksichtigt. Mit einem großen Teil dieser Datensets, die jeweils aus weit über einhundert Parametern (Variablen) für tausende von Patienten bestehen, wurde das Informations-System trainiert. Eine zusätzliche Anzahl Datensets von (knapp eintausend) zufällig gewählten Patienten, die nicht zum Trainieren des neuronalen Netzes eingesetzt wurden, wurde zur Prüfung des Systems benutzt.

[0014] Dieses System ermöglicht eine Voraussage der Überlebenschancen eines Patienten nach kurativer kolorektaler Chirurgie mit einem prädiktiven Wert von etwa 90%. Für die Metastatisierung kann das System diese gute prognostische Aussage noch nicht reproduzieren. Hier sind offensichtlich molekularbiologische Zusatzdaten erforderlich; um die Leistung zu verbessern. Am signifikantesten schneiden folgende Variablen ab: Tiefe der Tumordinfiltration, T-Kategorie, tumorfreie Resektionsränder, Grading, venöse und lymphatische Invasion.

[0015] Die Aufgabe, die durch die Erfindung gelöst werden soll, besteht darin, ein verbessertes Verfahren zur Datenauswertung bereitzustellen. Ziel ist die Erweiterung der Aussagefähigkeit (die bisher nur für Patientengruppen möglich war) auf Prognosen für einen individuellen Patienten (z. B. hinsichtlich Metastatisierungsrisiko, Therapieantwort auf eine Reihe von Chemotherapeutika, und Voraussage von Nebeneffekten). Des weiteren soll durch die Erfindung die Signifikanz der in der Auswertung berücksichtigten Parameter bestimmt und dadurch eine Reduzierung der für die Prognose notwendigen Datenmenge erreicht werden, ohne die Qualität der Prognose zu schmälern.

[0016] Diese Aufgabe wird erfindungsgemäß gelöst durch die Merkmale im kennzeichnenden Teil der Ansprüche 1 und 14 bis 17 im Zusammenwirken mit den Merkmalen im Oberbegriff. Zweckmäßige Ausgestaltungen der Erfindung sind in den Unteransprüchen enthalten.

[0017] Ein besonderer Vorteil der Erfindung liegt darin, daß bei dem Verfahren zur Datenauswertung einem Computersystem Erfahrungsdaten und objekt- oder personenspezifische Anfragedaten zugeführt werden, durch eine auf die Anfragedaten bezogene Anfrage diese Anfragedaten unter Benutzung der Erfahrungsdaten automatisch ausgewertet werden, die Ergebnisse der Auswertung automatisch bewertet werden, anhand dieser Bewertung, der zugehörigen Anfrage und den bei der Auswertung benutzten Daten automatisch ein Maß für den Einfluß dieser benutzten Daten auf die Auswertung ermittelt wird, die Ergebnisse der Auswertung, die Bewertung der Ergebnisse und/oder das Maß des Einflusses der benutzten Daten dem Anfragenden zur Verfügung gestellt werden.

[0018] Ein Computersystem zur Datenauswertung umfaßt Mittel zur Eingabe von Erfahrungsdaten und objekt- oder personenspezifische Anfragedaten, Mittel, durch welche diese Anfragedaten unter Benutzung der Erfahrungsdaten bei einer auf die Anfragedaten bezogene Anfrage automatisch ausgewertet werden, Mittel zur automatischen Bewertung der Ergebnisse dieser Auswertung, Mittel, durch welche anhand dieser Bewertung, der zugehörigen Anfrage und den bei der Auswertung benutzten Daten automatisch ein Maß für den Einfluß dieser benutzten Daten auf die Auswertung ermittelt wird, und Mittel zur Datenausgabe.

[0019] Die Erfindung umfaßt ein Computerprogrammprodukt zur Datenauswertung, das direkt in den internen Speicher eines digitalen Computers geladen werden kann und Softwarecodeabschnitte umfaßt, mit denen das Verfahren

gemäß Anspruch 1 ausgeführt wird, wenn das Produkt auf einem Computer läuft.

[0020] Ein weiterer Bestandteil der Erfindung ist ein Computerprogrammprodukt zur Datenauswertung, das auf einem computergeeigneten Medium gespeichert ist und folgendes umfaßt: computerlesbare Programmmittel, die ein Computersystem veranlassen, durch eine auf in dem Computersystem gespeicherte Anfragedaten bezogene Anfrage diese Anfragedaten unter Benutzung von in dem Computersystem gespeicherten Erfahrungsdaten automatisch auszuwerten, computerlesbare Programmmittel, die ein Computersystem veranlassen, die Ergebnisse der Auswertung automatisch zu bewerten, computerlesbare Programmmittel, die ein Computersystem veranlassen, anhand dieser Bewertung, der zugehörigen Anfrage und den bei der Auswertung benutzten Daten automatisch ein Maß für den Einfluß dieser benutzten Daten auf die Auswertung zu ermitteln.

[0021] Eine Methode zur Nutzung eines Systems zur Auswertung von objekt- oder personenspezifischen Daten besteht darin, daß der Anbieter dieses Systems die Eingabe von und/oder den Zugang zu objekt- oder personenspezifischen Datensätzen für Nutzer des Systems mit einer gebührenpflichtigen PIN koppelt derart, daß die PIN gleichzeitig mit einem Mittel zur Erfassung von in das System einzugebenden objekt- oder personenspezifischen Daten und/oder einem Mittel zur Erfassung von Material, welches zur Ermittlung dieser objekt- oder personenspezifischen Daten dient, verbunden ist, und der Nutzer die PIN zusammen mit diesem (diesen) Mittel(n) durch Zahlung einer Gebühr erwirbt.

[0022] Durch die Erfindung wird beispielsweise die Verfügbarkeit klinisch relevanten Wissens zu einem spezifischen medizinischen Problem am Arbeitsplatz des Arztes (Onkologen) erheblich verbessert. Über das (im Internet erreichbare) Computersystem erhält er Zugriff auf Daten, die aus verschiedenen Datenbanken stammen, aber z. B. durch den Einsatz des Internetstandards XML (eXtensible Markup Language) auf eine einheitliche Basis gestellt wurden. Für ihn sind die unterschiedlichen Formate und Datenstrukturen nicht mehr zu erkennen. Dadurch wird es möglich, die Akzeptanz und den tatsächlichen Einfluß von Genexpressionsdaten auf die Patientenversorgung nachhaltig zu beeinflussen. Durch diese Vereinheitlichung der Sprache bzw. Formate der Daten werden die organisatorischen Voraussetzungen für einen erfolgreichen Einsatz der Erfindung geschaffen.

[0023] Die Erfindung wird die Schätzung des Metastatisierungsrisiko eines individuellen Patienten erlauben, so daß die Indikation einer adjuvanten Chemotherapie gezielter gestellt werden kann. Außerdem wird eine Schätzung der Probabilität der Therapieantwort bei einer Reihe von Chemotherapeutika ermöglicht, so daß mit einer entsprechenden Tumorphilierung auch Tumoresistenzmuster erkannt werden können. Durch die Selektion der molekularen Targets (auf DNA, RNA und Proteine-Ebenen), die mit einem bestimmten klinischen Outcome (z. B. Metastatisierung) verbunden sind, wird eine ganz erhebliche Reduzierung der Daten erreicht, die für eine hochwertige Prognose ausgewertet werden müssen. Es werden nur die Outcome-relevanten Moleküle berücksichtigt, was einen entscheidenden Schritt in Richtung Validierung von potentiellen Drug-Targets im humanen Patienten darstellt. Als weiterer Vorteil der Erfindung ist anzusehen, daß das Risiko von erfolglosen teuren Therapieversuche deutlich gesenkt, die Entwicklungskosten eines Medikaments reduziert, und damit die Gesundheitskosten verringert werden, da man in der Lage ist, Patientenkollektive zu bestimmen, die für klinische Studien mit einem bestimmten Chemotherapeutikum am besten geeignet

sind. Mit Hilfe der Erfindung wird es möglich sein, zu beurteilen, ob eine zusätzliche Therapie zu einer deutlichen Verbesserung der Prognose führt, z. B. verglichen mit dem Pool der Patienten, die mit adjuvanten Therapie behandelt worden sind. Aufgrund dieser spezifischen Informationsmuster wird der Arzt in die Lage versetzt, eine individuelle – d. h. eine auf ein spezifisches Krankheitsbild bzw. Krankheitsstadium bezogene – Therapieentscheidung zu treffen.

[0024] So könnten möglicherweise durch den Einsatz der Erfindung prospektiv randomisierte Studien durch evidenzbasierte Daten mit einem hohen Wert ersetzt werden. Dies wäre ein weiterer Vorteil, denn die Durchführung von zahlreichen randomisierten Studien bei einer zunehmenden Zahl von Krebstherapien ist mit erheblichen Kosten und organisatorischen Schwierigkeiten verbunden, die somit eingespart werden könnten.

[0025] Erkenntnisse aus (Internet-)Management und Qualitätssicherung für e-Health-Systeme zeigen, daß die Zielsetzung, die durch die Erfindung umgesetzt wird, eine problemorientierte Darstellung erfordert, die sich nicht allein an einer streng wissenschaftlichen Gliederung orientiert, wie die Genexpressionsdatenvermittlung. Vielmehr ist es notwendig – und die Erfindung erfüllt diese Forderung –, mit Hilfe eines gut strukturierten Aufbaus, der Darstellung von größeren Zusammenhängen und dem Aufzeigen von Methoden und Techniken eine Grundlage zu schaffen, die es dem Arzt (Onkologen) ermöglicht, neben einer individuellen Therapieentscheidung außerdem eigene praktische Lösungsansätze auf Basis des vermittelten Wissens bilden zu können.

[0026] Die Erfindung soll nachstehend anhand von einem zumindest teilweise in den Figuren dargestellten Ausführungsbeispielen näher erläutert werden.

[0027] Es zeigen:

[0028] Fig. 1 Darstellung des modularen Aufbaus eines medizinischen Informationssystems,

[0029] Fig. 2 detaillierte Darstellung des modularen Aufbaus eines medizinischen Informationssystems.

[0030] Am Beispiel eines medizinischen Informationssystems für Onkologen, in welchem ein internetbasiertes patientenspezifisches Prognosesystem realisiert wurde, soll Einsatz und Wirkungsweise der Erfindung beschrieben werden.

[0031] Das beispielhafte System ist ein internetbasiertes medizinisches Informationssystem, das sich aus Datenbanken, einem Datenreduktionsprogramm und Modulen der künstlichen Intelligenz (neurales Netzwerk oder regelbasiertes System) zusammensetzt. Es erlaubt die Integration von klinischen, pathologischen und biologischen Daten, und deren Verknüpfung mit relevanten Prognoseaussagen für einen bestimmten Patienten. Damit erlaubt dieses Informationssystem dem Onkologen, aufgrund spezifischer Informationsmuster eine individuelle Therapieentscheidung zu treffen. Die Therapieentscheidungen werden mit Probabilitätsrechnungen unterstützt. Als Prototyp wurde das kolorektale Karzinom gewählt. Das beispielhafte medizinische Informationssystem integriert Daten aus der Transkriptom- und Proteom-Forschung.

[0032] Die Anwendung des beispielhaften Informationssystems wird im folgenden beschrieben. Ein Patient wird beim Arzt vorstellig und erkundigt sich nach Therapiemöglichkeiten für seine Krebserkrankung. Nach der Operation schickt der Onkologe die Proben gegen Rechnung an ein Referenzlabor, wo die notwendigen Genexpressionsanalysen durchgeführt werden. Mit Einsendung der Proben erhält der Onkologe eine PIN-Nummer, dank der er alle relevanten Daten des Patienten (patientenspezifisch, umweltspezifisch etc.) in anonymisierter Form in die Datenbank des erfin-

dungsgemäßen Computersystems eintragen kann. Unterdessen trägt das Referenzlabor ebenfalls alle mit der PIN versehenen Tumordaten in diese Datenbank ein. Zu Beginn der Nutzung umfaßt ein Datensatz alle Prognosefaktoren als Variablen, die von der Medizin für die Prognose bei kolorektale Karzinom akzeptiert werden. Mit dem erfindungsgemäßen Verfahren kann die Korrespondenz der molekularen und klinischen Informationen erkannt werden. Diese kombinierten Informationen werden mit der Datenbank verglichen, und die Patienten mit dem nächstliegenden Informationsmuster und deren Krankheitsverlauf wird herausgesucht. Es wird dabei ein rückwirkendes Fehlerminimierungsverfahren angewendet. Der Arzt (Chirurg oder Onkologe) kann anschließend mit der PIN beim Informationssystem anfragen. Das System fragt nach klinisch-pathologischen Daten, die der Arzt anonymisiert eingibt. Innerhalb von Minuten erhält dieser dann Angaben zu Metastatisierungswahrscheinlichkeit, Resistenzprofile zu verschiedenen Chemotherapeutika, möglicherweise zu Immuntherapeutika oder ähnlichem. Die so gewonnenen Erkenntnisse bilden damit eine wichtige Entscheidungshilfe bei der Therapieentscheidung. Patienten erhalten keinen direkten Zugriff auf die Webseite, die den Zugang zu dem erfindungsgemäßen Computersystem gestattet. Später wird der Arzt von der Datenbank regelmäßig über Therapie und Verlauf seines Patienten befragt, diese Daten werden zum iterativen Lernprozeß des Systems gebraucht, so daß sich dieses laufend dem medizinischen Fortschritt anpassen kann. Dadurch werden signifikante und unwesentliche Variablen ermittelt, was zu einer Optimierung der auszuwertenden Datenmenge und zur Verbesserung der Prognose führt.

[0033] Zum Verständnis der Herkunft und Verknüpfung von Daten muß verdeutlicht werden, daß die heute verfügbaren signifikantesten Daten aus der Klinik, der Pathologie, und der gegebenenfalls schon erfolgten Behandlung es erlauben, eine prognostische Aussage zu treffen. Diese Prognostik wurde im beispielhaften Informationssystem für das kolorektale Karzinom durch moderne Bio-Informatik optimiert. Dieses System hat eine prognostische Leistung erreicht, die bei Hunderten von Patienten durch Kreuzvalidierung genau bestimmt werden konnte. Die Einbeziehung von neuen (molekularbiologischen) Daten erlaubt es, das System erneut zu "trainieren". Wenn die prognostische Leistung steigt, werden die neuen Daten als prognostisch Signifikant bewertet und für weitere Analysen erfordert. Falls das System mit diesem Set neuer Daten nicht besser wird, werden diese Daten ausgeschieden. So können biologische Daten äußerst effizient ausgewählt werden. Mit dem Molecular-Mining System z. B. steht ein Auswahl-Verfahren zur Verfügung, welches dies ermöglicht.

[0034] Im Gegensatz zu klinischen Versuchen, gibt es keine Standards für Prognosefaktor-Studien. Fast alle Prognosefaktor-Studien haben leider eher einen erklärenden als einen beweisenden Charakter. Es ist daher wichtig für klinische Forscher, Standards für die Beweise für einen Prognosefaktor zu definieren, bevor er in der Praxis angewendet wird. Folgende Richtlinien sollten gelten und liegen dem erfinderischen System zugrunde:

- die Reproduzierbarkeit der Untersuchung im eigenen und in anderen Laboren,
- Untersuchung ungeachtet des Ergebnisses (assay blinded to outcome)
- es sollten von weniger als 15% der Patienten Daten fehlen,
- einheitliche Behandlung,
- Hypothese im voraus,
- genügend Patienten (> 10 pro Ereignis),

- durch neue Faktoren wird das derzeitige Wissen durch Vorhersagen ergänzt,
- angepaßte Analysen für die verschiedenen Hypothesen,
- Studienabgrenzungen müssen vorher spezifiziert werden.

[0035] Diese Richtlinien sind nicht nur für Studien sondern auch für den Erfolg eines Prognosefaktors von Bedeutung. Sie wurden in dem erfinderischen System berücksichtigt. Die Anerkennung eines neuen Faktors kann nur erfolgreich sein, wenn mindestens eine belegende Studie vorliegt und wenn die Untersuchungen in mehreren Kliniken reproduzierbar sind. Der Prognosewert sollte über bisherige Standard-Prognosefaktoren hinaus gehen, und er muß Auswirkungen auf die Therapie haben.

[0036] Um eine Prognose zu bestimmen, sollte man von drei Prognosefaktoren ausgehen.

- Tumorbezogene Faktoren: charakterisieren die Krankheit,
- patientenspezifische Faktoren: beziehen sich auf den Patienten,
- umweltspezifische Faktoren, die sich weder direkt auf den Patienten noch auf den Tumor beziehen.

[0037] Dabei sollten vorteilhafterweise die folgenden Punkte berücksichtigt werden, wobei diese erforderlichenfalls gemäß neuen Erkenntnissen ergänzt werden können.

Tumorspezifische Faktoren

[0038] Diese Faktoren sind eigentlich immer die Determinanten, die bestimmenden Faktoren für das Ergebnis in Krebspatienten. Die wichtigsten tumorspezifischen Faktoren beziehen sich auf histologische Angaben (Typ, Merkmale) und die anatomische Ausbreitung der Krankheit.

Pathologie des Tumors

[0039] Die Tumorpathologie ist entscheidend für die Prognose bei Krebs. Der histologische Typ definiert zwar die Krankheit, aber andere Faktoren, wie z. B. das Stadium oder der Befall von Lymphknoten beeinflussen das Ergebnis.

Ausbreitung der Krankheit

[0040] Die anatomische Ausbreitung des Tumors wird gewöhnlich nach den Kriterien der TNM-Klassifikation zu Größe, Infiltration des Primärtumors, vorhandenen Lymphknotenmetastasen und Fernmetastasen beschrieben.

Tumorbiologie

[0041] Bisher wurden krebspezifische Proteine lediglich als Tumormarker verwendet, um die Tumorerlastung wiederzuspiegeln, ohne jedoch das Tumorverhalten genau charakterisieren zu können. Jüngere Ergebnisse in der Tumorbiologie haben die prognostische Rolle von tumorspezifischen Proteinen wieder in den Vordergrund rücken lassen. Als Genprodukte können sie u. a. Ursachen und Suppression von Krebs, die normale und anormale Kontrolle des Zellzyklus, und Metastasierung und Angiogenese des Tumors bestimmen. Neue Technologien in der molekularen Diagnostik machen es heute möglich, genetische Information bezogen auf minimale Tumorerlastung, aggressives Tumorzellwachstum und Tumorzellreaktion infolge von Veränderungen der DNA oder Immuntherapien zu bestimmen.

men.

Tumorspezifische Symptome

[0042] Obwohl sie auch als patientenspezifisch angesehen werden können, ist die eigentliche Ursache für Symptome in der Onkologie der invasive Charakter des Tumors. In der Tat sind Symptome bei den meisten Krebspatienten ein sehr wichtiger Prognosefaktor. Klassische Beispiele für die Wirkung von Symptomen sind die B-Symptome (Nachtschweiß, Fieber und Gewichtsverlust).

Patientenspezifische Faktoren

[0043] Dies sind im Patienten vorhandene Faktoren, die entweder indirekt oder gar nicht maligne sind, die jedoch durch eine Interferenz mit dem Tumorverhalten oder ihre Reaktion auf die Behandlung einen großen Einfluß auf das Ergebnis haben können. Hier unterscheidet man zwischen demografischen Faktoren, Komorbidität und zugleich existierenden Krankheiten.

Demografische Faktoren

[0044] Diese Faktoren, die sich auf das onkologische Ergebnis auswirken, sind Alter, Geschlecht und ethnische Zugehörigkeit. Keiner dieser Faktoren kann durch einen Eingriff bzw. eine Behandlung beeinflusst werden, aber viele beeinflussen unabhängig voneinander das Ergebnis anderer Faktoren. Beispielsweise haben ältere Patienten eine niedrigere Überlebenszeit bei der Hodgkin'schen Krankheit oder bei Lymphomen. Die Rolle des Geschlechts ist weitaus weniger genau definiert, aber bei Hodgkin Erkrankungen oder bösartigen Malignomas waren die Ergebnisse bei Männern schlechter als bei Frauen.

Komorbidität

[0045] Diese Faktoren können vererbte genetische Krankheiten, wie z. B. Neurofibromatose sein, die ein Risikofaktor für neurogene Sarkomas und einen prognostischen Faktor für Krebsergebnisse darstellen.

Performance Status

[0046] Der Performance Status ist ein starker prognostischer Faktor für viele Krebsarten, besonders bei denen im fortgeschrittenen Status wie z. B. Lungen- und Blasenkrebs, die eine Chemotherapie erfordern. Als Folge des Alters oder der Komorbidität sollte dieser Faktor als patientenspezifischer Faktor angesehen werden.

Übereinstimmung

[0047] Übereinstimmung mit einer vorgeschlagenen Krebsvorsorgeuntersuchung oder Behandlungsplan kann das Überleben eines Patienten oder einer Gruppe von Patienten beeinflussen. Fehlende Übereinstimmung mit Krebsvorsorgeempfehlungen bei Brustkrebs kann zu einer verspäteten Diagnose, einem weiter fortgeschrittenen Stadium bei der Diagnose und schlechterem Überleben führen.

Umweltspezifische Faktoren

[0048] Obwohl die umweltspezifischen Faktoren weniger untersucht wurden und oft nicht in die Diskussionen einbezogen werden, haben sie einen Einfluß auf das Ergebnis für einen einzelnen oder eine ganze Gruppe von Patienten.

Mediziner

[0049] Der Behandlungsplan hat eine tiefgreifende Wirkung auf das Ergebnis. Unangemessene Eingriffe können in exzessiver Toxizität und eingeschränkter Lebensqualität enden. Fehlgeschlagene Kontrolle des Krebses kann auch den Tod für den Patienten bedeuten. Die Expertise des behandelnden Arztes ist ein weiterer prognostischer Faktor, da sie ebenfalls das Ergebnis im Krebspatienten beeinflusst. Es gibt zunehmende Beweise dafür, daß Kliniken, die keine bestimmte Anzahl ("critical mass") an Patienten behandelt, auch keine optimalen Behandlungsergebnisse erreichen.

Gesundheitswesen

[0050] Hier gibt es große Divergenzen zwischen einzelnen Bevölkerungsgruppen. Mehrere Studien haben belegt, daß z. B. ältere Menschen (75 Jahre und älter) oder Patienten aus anderen ethnischen Gruppen nicht die gleiche Behandlung bekommen wie z. B. jüngere oder einheimische Patienten und somit ihr Behandlungsergebnis beeinflusst wird.

Gesellschaft

[0051] Studien des Office of National Statistics (GB) haben gezeigt, daß das Überleben bei einer Krebserkrankung abhängig ist vom sozial-wirtschaftlichen Stand der Patienten. Ein weiterer Faktor für eine schlechtere Prognose ist die Ernährung.

[0052] Für die Genexpressionsprofilierung auf RNA- und Protein-Ebene werden purifizierte Proben und klinische Daten von kolorektalen Patienten, von denen die erforderlichen Angaben zur Verfügung stehen, analysiert und daraus die Transkriptom- und Proteom-Profile bestimmt. Hierzu steht eine Vielzahl von tiefgekühlten Proben von verschiedenen kolorektalen Patienten aus verschiedenen Institutionen zu Verfügung.

[0053] Diese Proben wurden speziell für diesen Zweck entnommen, und sind charakterisiert durch eine ausgezeichnete Qualität und Reproduzierbarkeit. Der Anteil an Epithelzellen variiert erheblich zwischen verschiedenen Präparaten (Reymond et al. Electrophoresis 1997a). Für die Probenvorbereitung wurde eine Methode entwickelt, die es erlaubt, reine Epithelzellen in genügender Menge (über 10⁸ Zellen) aus chirurgischen Präparaten vorzubereiten (diese Methode ist z. B. beschrieben in UK Patent Application GB 9705949.7). Aus diesen Proben konnten sowohl Proteine, als auch RNA vorbereitet werden, die qualitativ mit Produkten aus Zelllinien verglichen werden können.

[0054] Die nach dieser Methode vorbereiteten Proben können verglichen werden, auch wenn sie aus verschiedenen Institutionen kommen. Damit ist eine Grundbedingung erfüllt, um später die prädiktiven Aussagen des hier beispielhaft beschriebenen Systems in verschiedenen Institutionen zu vergleichen. Damit können in der Zukunft die großen Probendurchsatzzahlen erreicht werden, die zur Validierung der Genexpressionsforschung notwendig sind. Bis heute wurden mehrere Tausend Proben von Patienten, z. T. mit Stuhl, Blut und Knochenmarkpunktion, gewonnen. Der theoretische Input des klinischen Netzwerkes, in dem das beispielhafte Informationssystem integriert ist, beträgt mehrere tausend neue kolorektale Karzinome pro Jahr.

[0055] Diese Proben sind mit klinisch-pathologischen Daten verbunden, die an die Anforderungen des beispielhaften medizinischen Informationssystems angepaßt sind, d. h. daß alle Parameter (Variablen) in jeder Institution gesammelt worden sind. Das (gemeinsame) Follow-up Schema ent-

spricht den deutschen Richtlinien.

[0056] Die persönlichen Patientendaten werden in den teilnehmenden Institutionen behalten, die das Follow-up der Patienten sichern. Nur anonymisierte Daten werden an das Informationssystem weitergegeben.

[0057] Zur Durchführung einer qualitativ hochwertigen Genexpressionsanalyse steht inzwischen vielfältige Technik zur Verfügung, mit der man in der Lage ist, das Expressionsniveau jedes bekannten Gens auf Transkriptom- und Proteomebene zu analysieren. Ein Komplettsystem für derartige Analysen umfaßt beispielsweise die folgenden Komponenten:

Transkriptom-Analyse

- Die DNA-Chips (DNA-Mikroarrays): Man unterscheidet grundsätzlich die cDNA-Chips von den Oligo-Chips. Bei cDNA-Chips werden ca. 300-400 bp lange PCR Produkte auf die Chips aufgebracht. Man ist derzeit in der Lage ca. 14.000 cDNAs auf ein Array zu spotten.

Bei den Oligo-Chips werden ca. 60 bp lange Oligonukleotide auf der Chip-Oberfläche synthetisiert. Man produziert Arrays mit 8.400 Features, auf Wunsch auch als Doppelarray (16.800 Spots). Gerade im Bereich der DNA-Chips werden Neuentwicklungen in sehr kurzer Zeit zu immer dichteren Arrays (höhere Spotanzahl) mit einer sehr großen Flexibilität bei der Sequenzauswahl führen.

- Der Mikroarray-Scanner: Neuentwicklungen auf dem Gebiet der Mikroarray-Scanner sind in der Lage gleichzeitig 2 Fluoreszenzwellenlängen zu analysieren. Bei einer Auflösung von 5 oder 10 µm (vom Benutzer einstellbar) benötigt der Scanner ca. 8 min für das scannen eines Chips. Ein 48 Chips fassendes Karussell erlaubt die Benutzung dieses Systems in der Hochdurchsatz-Analyse.

- Der Bioanalytiker: Der Bioanalytiker ist ein Lab-on-a-Chip-System, das zur Qualitätskontrolle vor allem bei der RNA-Aufreinigung dient. Mit Hilfe des Bioanalyzers wird die vom Experimentator aufgereinigte RNA maschinell qualitativ und quantitativ analysiert.

Proteom-Analyse

[0058] Mit Proteom-Techniken kann die qualitative und quantitative Expression von Proteinen in verschiedenen Krankheitsstadien bestimmt werden. Da bekannterweise post-translationelle Proteinänderungen eine wichtige Rolle im klinischen Verhalten kranker Zellen, Gewebe und/oder Organe bedeuten, haben diese Unterschiede in der Proteinexpression einen wichtigen Einfluss in der Anwendung des beispielhaft beschriebenen Informationssystems.

[0059] Als Proteom-Techniken im menschlichen kolorektalen Karzinom können z. B. ein- (SDS-PAGE) oder zweidimensionale Gel-Elektrophorese (2D PAGE), N-terminale Sequenzierung und Massenspektrometrie (MALDI-TOF und MS-MS) sowie Chips, auf denen Antikörper-, Ligand- oder verschiedene Oberflächen zur Bindung von Proteinen angewendet werden, dienen.

[0060] Als beispielhafte Technologie für Arbeiten auf dem Gebiet der Proteom-Forschung bietet sich eine Kombination der beiden Schlüsseltechnologien 2-D-Gelelektrophorese und Massenspektrometrie an. Die Proteine werden mittels der 2-D-Gelelektrophorese aufgetrennt und anschließend gefärbt. Die Protein-Spots werden ausgeschnitten, enzymatisch verdaut und das entstehende Peptidgemisch massenspektrometrisch untersucht. Über einen Datenbankab-

gleich der resultierenden Peptide Mass Fingerprints erfolgt die Proteinidentifizierung.

[0061] Die Massenspektrometrieplattform setzt sich in diesem Falle aus einer automatischen Probenvorbereitungsstation, einem High-Performance MALDI-Massenspektrometer (Matrix-Assisted Laser Desorption/Ionization Time of Flight) und einer Datenstation zur automatischen Durchführung der Datenbanksuche zusammen. Das MALDI-MS verfügt über eine hohe Sensitivität und hohe Massengenauigkeit; beides sind Grundvoraussetzungen für eine erfolgreiche Proteinidentifizierung. Zusätzlich können mittels der PSD-Technik (Post Source Decay) Sequenzinformationen der Peptide ermittelt werden.

[0062] Um einen leichteren Zugang zu Sequenzinformationen zu bekommen und gezielt posttranslationale Modifikationen bestimmen zu können, ist der Einsatz eines Elektrospray-Massenspektrometers hilfreich. Durch den Einsatz eines automatischen Spotpickers und eines Digesters kann eine weitere Automatisierung erreicht werden.

[0063] Eine weitere wichtige Aufgabe, die bei der Entwicklung des erfindungsgemäßen medizinischen Informationssystems gelöst werden mußte, ist die Übersetzung der verschiedenen Informations-Plattformen (Transkriptom- und/oder Proteomdaten) in eine gemeinsame Sprache.

[0064] Um dieses Konvertierungsproblem zu lösen, wurde für das beispielhafte onkologische Informationssystem ein Bioinformatik-Konzept entwickelt, daß es erlaubt, Daten aus der Klinik, aus der Pathologie, aus DNA-Databasen (wie z. B. CGAP), aus cDNA-Arrays (wie z. B. Agilent Chips) und aus 2D PAGE zu integrieren und zu analysieren. Die verschiedenen Informationen aus der Klinik, der Pathologie, der Transkriptom- und der Proteom-Forschung werden in die Web-basierte (*.xml) bioinformatische Sprache GEML (Gene Expression Markup Language) übersetzt (s. <http://www.geml.org>).

[0065] Damit wurden die Voraussetzungen für ein echtes "Bridging" der verschiedenen Informationsbasen geschaffen. Dieses "Bridging" ist als eine unabdingbare Voraussetzung für die Umsetzung der experimentellen Ergebnisse der Gen-Expression in eine klinisch sinnvolle Information zu betrachten.

[0066] Nachdem nun die Daten aus Klinik und Labor in das *.xml Format übersetzt wurden, kann man auf eine (vereinheitlichte) Datenbank zugreifen, die etwa 10^4 Informationen enthält. Diese etwa 10^4 Informationen, die jetzt für jeden Patienten zur Verfügung stehen, können keinesfalls dem Onkologen direkt vermittelt werden. Aus diesem Grunde muß in das erfindungsgemäße Informationssystem eine Software integriert sein, die eine Auswertung dieser Fülle an Information unterstützt.

[0067] Als Komponente des beispielhaften onkologischen Informationssystems findet ein Modul mit Datenreduktions-Software Verwendung. Dabei musste verfügbare Datenreduktions-Software an die speziellen Anforderungen der onkologischen Auswertungen angepasst werden. Die etwa 10^4 Informationen pro Patient werden durch den Einsatz dieser Software auf 10^2 reduziert.

[0068] Die vom Scanner erzeugten digitalisierten Proteom- oder Transcriptom-Bilder werden in einem kompatiblen Analyseprogramm verarbeitet. Dieses Programm ist, in der Lage die Genexpressionsdaten auszuwerten und zu speichern. Das Programm protokolliert jedes Genexpressionsmuster und erlaubt Vergleiche verschiedener Experimente. Hierzu sind i. a. Datenbankabfragen in externen sowie internen Datenbanken notwendig. Das Programm generiert technologiespezifische Irrtum-Modelle. Die aus den Irrtum-Modellen generierten Probabilitätswerte jeder Messung werden über die gesamte Analyse-Umgebung propagiert, was einen

höheren prädiktiven Wert bei Cluster-Analysen, Ähnlichkeiten-Suchen und Tendenz-Analysen ermöglicht. Durch besondere informationstechnische Werkzeuge ermöglicht das Programm, Analysen über Exons, Sequenzen, Cluster Intensität- und Verhältnisrechnungen durchzuführen. Clustering-Analysen beinhalten beispielweise agglomerative, Teilungs-, Mittelwerte und Medianwerte-Algorithmen. Ein beispielhaftes informationstechnisches Verfahren erlaubt es, ähnliche Muster mit dem Muster von Interesse innerhalb aller Datensets der Datenbank zu recherchieren. Auch Zeit-Sequenzen, beispielhaft in einer iterativen Nachsorgemessung, können über eine Zeitlinie dargestellt werden, wodurch spezifische Verhalten identifiziert werden können. Besondere Suchmaschinen erlauben eine rasche Datenbankanfrage, und können an eine interne Datenbank angepaßt werden. Auch Hypertext-links können formuliert werden, so daß Verbindungen mit internen oder externen Datenbanken hergestellt werden können.

[0069] Durch dieses Vorgehen wird erreicht, daß nur die biologischen Daten, die für eine bestimmte klinische Beobachtung ein signifikantes Verhalten zeigen, berücksichtigt und für die Auswertung übernommen werden. Dadurch wird gegenüber herkömmlichen medizinischen Informationssystemen eine erhebliche Datenreduktion erreicht.

[0070] Die Bioinformatik wird weiterhin dadurch unterstützt, daß klinische Outcomes (wie die Metastatisierungsfähigkeit oder die Therapie-Resistenz eines bestimmten Tumors) nach Berücksichtigung der klinisch-pathologischen Daten direkt mit Datenmustern verbunden werden können. Diese Interpretation kann z. B. durch künstliche Intelligenz und/oder maschinelle Lernverfahren vereinfacht werden. Herkömmliche Computerprogramme umfassen eine Menge von expliziten Anweisungen, die dem Programm genau sagen, was und wie es eine Berechnung auszuführen hat. Systeme der künstlichen Intelligenz (KI-Systeme) arbeiten unter vollkommen anderen Voraussetzungen: dem Programm werden eher Kenntnisse vermittelt, als daß ihm exakte Anweisungen für die Abarbeitung gegeben werden. Dies passiert während der Trainingsphase des KI-Systems. Indem das KI-System wiederholt auf historische Daten angewandt wird und die Ergebnisse dieser Auswertungen ("Schlussfolgerungen") mit den tatsächlich vorliegenden Fakten verglichen werden, lernt es im Verlaufe dieses Trainings das Verhalten, das von dem "fertig ausgebildeten" System gefordert wird.

[0071] Der Korrespondenzanalyseansatz und die aufsteigend hierarchische Klassifikation, die bei dem erfindungsgemäßen Informationssystem angewendet werden, weichen signifikant von dem eher klassischen Ansatz der Diskriminanzanalyse mittels Hauptkomponentenanalyse ab. Beginnend mit einer Reihe von Experimenten, wobei jedes Experiment eine große Anzahl an Datenpunkten aufweist, liefert die Korrespondenzanalyse für die Repräsentation der Proben einen faktorisierten Raum von reduzierter Größe. Die ansteigende hierarchische Klassifikation sortiert die Bilder in aussagefähige Gruppen. Die simultane Darstellung sowohl der Spots als auch der Chip- oder Gel-Bilder findet in dem selben faktorisierten Raum statt. Die charakteristischen Gen- oder Protein-Vertreter einer bestimmten Klasse von Gelen (z. B. Krebsmetastasen-Proben) werden präzise markiert; was die Analyse erheblich vereinfacht. Dementsprechend kann die Software automatisch Protein- oder Gen-Muster klassifizieren, wobei dafür den jeweiligen Anforderungen entsprechend Hauptkomponentenanalyse, Wavelet-Analyse, künstliche neuronale Netze, heuristischer Clusterbildungsanalyse und andere einzeln oder in Kombination eingesetzt werden können.

[0072] Um die mit der Software ausgewählten Parameter

flächendeckend und billig analysieren zu können, wird ein reproduzierbarer Labortest durchgeführt.

[0073] Dieser Labortest erlaubt es, Outcome-relevante genetische, translationelle oder funktionelle Charakteristiken eines Tumors zusammenzufassen. Dieses Vorgehen ermöglicht es, sogenannte Integrated Health-Care Solutions anzuwenden, wo die Therapie mit der Diagnostik gekoppelt ist.

[0074] Die zu dem reproduzierbaren Labortest gehörigen Chips weisen folgende Eigenschaften auf:

- der Chip liefert Daten zum Metastatisierungspotential,
- der Chip liefert Daten zur Therapieantwort bei mindestens 10 verbreiteten Chemotherapeutika,
- der Chip liefert Daten zum Patientenmetabolismus (z. B. enzymatische Ausrüstung),
- der Chip liefert Information zur Autoimmunität gegen das Tumor,
- der Chip beinhaltet nicht mehr als 10^2 verschiedene Informationen (+ doubles) und
- der Chip ist reproduzierbar.

[0075] Durch die Reproduzierbarkeit wird eine breite Anwendbarkeit der Chips erreicht, so daß eine preiswerte Herstellung möglich wird.

[0076] Da die relevanten biologische Daten sich je nach Diagnose unterscheiden, muß für jede Diagnose ein eigener Test entwickelt werden. Die daraus resultierenden Labortests können sich dabei erheblich von denen unterscheiden, die beim kolorektalen Karzinom Outcomerelevant sind. Dadurch ist es schwierig, dieses Labortest genauer zu beschreiben.

[0077] Für den Datenaustausch zwischen behandelndem Arzt, Referenzlabor(en) und Datenbanken umfaßt das erfindungsgemäße Informationssystem eine vorzugsweise mehrsprachige, sichere Web-Schnittstellen, die bei der beispielhaften Lösung die Verbindung mit Onkologen und Referenzlabors ermöglichen. Mit der Strukturierung des Informationssystems mittels des Internetstandards XML (eXtensible Markup Language) verbessert sich die Verfügbarkeit klinisch relevanten Wissens zu einem spezifischen medizinischen Problem am Arbeitsplatz des Onkologen.

[0078] Die gewählte kryptographische Basistechnik des beispielhaften onkologischen Informationssystems ist die symmetrische Verschlüsselung. Hier stehen hocheffiziente Verfahren zur Verfügung, die bei einer Schlüssellänge von beispielsweise 128 bits langfristige Sicherheit gewähren. Kommunikationspartner verfügen über einen gemeinsamen Schlüssel, die PIN-Nummer. Die PIN-Nummer wird nur dem behandelten Arzt gegen Rechnung zugestellt, so daß der Patient keinen direkten Zugang zu den Aussagen des Informationssystems bekommen kann. Als vorteilhafter Standard kann beispielsweise der AES (Advanced Encryption Standard) gewählt werden.

[0079] Als sicherer Aufbewahrungsort für die elektronische Identität eines bestimmten Onkologen ist die kryptographische Chipkarte ideal, also im Gesundheitswesen die HPC (Health Professional Card, Ausweis für Berufe im Gesundheitswesen).

[0080] Zunächst wird eine Plattform definiert. Das heißt Inputs und Outputs werden festgelegt. Hierzu werden Festlegungen zum Informationsfluß und zum Abrechnungsmodell getroffen.

[0081] Für diesen Informationsfluß wird beispielhaft der HON Code of Conduct (HONcode) für medizinische Webseiten im Gesundheitsbereich umgesetzt (www.hon.ch/HONcode/German). Auf aktive Inhalte wie Java-Skripte wird verzichtet, außer in zwingenden Anwendungsfällen

wie der Fern-Eingabe von mit PIN anonymisierten klinischen und pathologischen Daten und von Nachsorgedaten.

[0082] Konkrete Maßnahmen sind auch für die Sicherheit des Servers des beispielhaften Informationssystem getroffen worden. Nur die nötigsten TCP-Dienste laufen auf dem Server. Der Mail-Server wird mit aktuellen Virusfiltern ausgestattet. Client/Server-Verbindungen, auch mit anonymisierten Daten, werden über SSL (Secure Socket Layer) abgewickelt. Für die Nutzer-Authentisierung werden beispielhaft X.509 Zertifikate eingesetzt. Schließlich wird die Datenbank des beispielhaften Informationssystems regelmäßig gesichert, um nach einem Schadenfall wieder einen sauberen Zustand herstellen zu können.

[0083] Der Betreiber dieses Informationssystems kann zur kommerziellen Nutzung der Erfindung entweder selbst einen Vertrieb aufbauen oder beispielsweise Referenzlabors, große Pharma-Unternehmen oder Internet Content Provider mit dem Vertrieb beauftragen. Diese wiederum könnten das Informationssystem dem Zielpublikum zur Nutzung anbieten. Die Referenzlabors, großen Pharma-Unternehmen oder Content Provider stellen dabei die eigentlichen Kunden des Betreibers ("Kunden") dar; das Zielpublikum bzw. Nutzer des Informationssystems ("Teilnehmer") sind z. B. Ärzte, die sich mit Krebserkrankungen beschäftigen.

[0084] Dabei würden durch diese Geschäftsidee mehrere Vorteile genutzt. So müßte sich der Betreiber nur auf wenige Kunden konzentrieren und könnte z. B. das entwickelte Vertriebssystem dieser großen Kunden nutzen, das bei den großen Pharma-Unternehmen bis zu dem einzelnen Arzt reicht. Durch die geeignete Wahl der Kunden kann die weltweite Verfügbarkeit des Informationssystems erreicht werden.

[0085] Die Abrechnung der Gebühren für die Nutzung des Informationssystems sollte vorteilhafterweise ebenfalls über die Kunden und nicht direkt über die einzelnen Teilnehmer aus dem Zielpublikum erfolgen. Dabei könnte man je nach Erfordernis eine Vorauszahlung, Ratenzahlung, Gebühren pro Nutzungsfall oder Gebühren prozentual zum Umsatz, den der Kunde mit dem Informationssystem macht, vereinbaren. Auch Gebühren je "PIN", die an die Teilnehmer des Zielpublikums oder Kunden vergeben wurden, sind denkbar. Diese PIN ermöglicht dem Teilnehmer aus dem Zielpublikum den Zugang zum Informationssystem und dessen Nutzung; genauer: zu den mit der PIN verknüpften patientenspezifischen Daten.

[0086] Der Teilnehmer erhält die PIN gegen Bezahlung einer Gebühr an den Kunden ("no money - no PIN"). In Verbindung mit der PIN erhält der Teilnehmer einen Chip, auf welchem die für die Analyse erforderlichen Tests enthalten sind. Die Anforderungen an den Chip (Art und Anzahl der enthaltenen Tests) ergeben sich u. a. auch aus den Aussagen, die das Informationssystem hinsichtlich der Signifikanz der Variablen macht. Nach Aufbringung der Patientenprobe wird der Chip dann an ein Referenzlabor geschickt, und die Auswertung erfolgt in der oben beschriebenen Weise. Ebenso ist es denkbar, daß die Patientenprobe vom Teilnehmer direkt an das Referenzlabor geschickt wird und erst dort auf den Chip aufgebracht werden. Die PIN würde dann vom Referenzlabor gemeinsam mit den Testergebnissen dem Teilnehmer übersandt. Der Preis für die Nutzung des Informationssystems wäre im Gesamtpreis für den Kauf des Chips enthalten.

[0087] Wenn der Teilnehmer aus dem Zielpublikum eine Anfrage an das Informationssystem stellt, wird er in der Regel aufgefordert, gewisse Angaben zu machen, insbesondere über Verlauf der Therapie, Medikamentierung oder Verlauf der Erkrankung. Diese Daten dienen u. a. dazu, das System zu optimieren. Da damit der Wert des Systems gesteigert wird, kann man von dieser Dateneingabe möglicherweise

auch eine Rückerstattung eines gewissen Betrages der Gebühren an den jeweiligen Teilnehmer erwägen.

[0088] Die Erfindung ist nicht beschränkt auf die hier dargestellten Ausführungsbeispiele. Vielmehr ist es möglich, durch Kombination und Modifikation der genannten Mittel und Merkmale weitere Ausführungsvarianten zu realisieren, ohne den Rahmen der Erfindung zu verlassen.

Patentansprüche

1. Verfahren zur Datenauswertung, **dadurch gekennzeichnet**, daß
 einem Computersystem Erfahrungsdaten und objekt- oder personenspezifische Anfragedaten zugeführt werden,
 durch eine auf die Anfragedaten bezogene Anfrage diese Anfragedaten unter Benutzung der Erfahrungsdaten automatisch ausgewertet werden,
 die Ergebnisse der Auswertung automatisch bewertet werden,
 anhand dieser Bewertung, der zugehörigen Anfrage und den bei der Auswertung benutzten Daten automatisch ein Maß für den Einfluß dieser benutzten Daten auf die Auswertung ermittelt wird,
 die Ergebnisse der Auswertung, die Bewertung der Ergebnisse und/oder das Maß des Einflusses der benutzten Daten dem Anfragenden zur Verfügung gestellt werden.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Datenauswertung medizinische Prognosen generiert und die objekt- oder personenspezifischen Anfragedaten als klinisch-pathologische Daten durch einen Arzt und/oder als biomolekulare Daten durch ein Analyselabor dem Computersystem zugeführt werden.
3. Verfahren nach einem der Ansprüche 1 oder 2, dadurch gekennzeichnet, daß die objekt- oder personenspezifischen Anfragedaten und/oder die Erfahrungsdaten
 tumorbezogene Faktoren und/oder
 patientenspezifische Faktoren und/oder
 umweltspezifische Faktoren
 berücksichtigen.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß eine Aktualisierung der Menge der Erfahrungsdaten erfolgt, indem Daten über die Therapie und den Krankheitsverlauf zu von dem Computersystem prognostizierten Fällen dem Computersystem zugeführt werden.
5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß eine Aktualisierung der Bewertungsvorschriften in einem iterativen Lernprozeß erfolgt, in welchem Anfrage, Anfragedaten, bei der Auswertung der Anfrage benutzte Erfahrungsdaten, Ergebnisse der Auswertung und tatsächlich eingetretene Ereignisse berücksichtigt werden.
6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß in Abhängigkeit von der Bewertung die Menge der Erfahrungsdaten reduziert wird.
7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß die Anfragedaten einen aus mindestens einem Parameter bestehenden Datensatz bilden.
8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß in Abhängigkeit von der Bewertung die Zahl der Parameter reduziert wird.
9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß die Auswertung durch

- Cluster-Analyse und/oder
 Ähnlichkeiten-Suche und/oder
 Tendenz-Analyse und/oder
 Korrespondenzanalyse und/oder
 aufsteigend hierarchische Klassifikation und/oder
 Hauptanalyse und/oder
 Wavelet-Analyse
 in Verbindung mit aus Irrtums-Modellen generierten
 Probabilitätswerten und/oder Mitteln der künstlichen
 Intelligenz wie
 neuronalen Netzen und/oder
 regelbasierten Systemen
 erfolgt.
10. Verfahren nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß die Bewertung der Ergebnisse erfolgt, indem die Eintrittswahrscheinlichkeit des durch die Anfrage bezeichneten Ereignisses aus den Erfahrungsdaten ermittelt wird.
11. Verfahren nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß das Maß des Einflusses von Daten derart ermittelt wird, daß eine Anfrage ein erstes Mal ohne Berücksichtigung dieser Daten und ein zweites Mal unter Berücksichtigung dieser Daten ausgewertet wird, die Bewertungen der Ergebnisse dieser beiden Auswertungen miteinander verglichen und das Maß des Einflusses der Daten relativ zu einer Verbesserung bzw. Verschlechterung der zweiten Bewertung gegenüber ersten Bewertung bestimmt wird.
12. Verfahren nach Anspruch 11, dadurch gekennzeichnet, daß
 bei einer Verbesserung der zweiten Bewertung gegenüber ersten Bewertung die Daten als signifikant angesehen und bei künftigen Auswertungen berücksichtigt werden,
 bei einer Verschlechterung der zweiten Bewertung gegenüber ersten Bewertung die Daten als nicht signifikant angesehen und für künftige Auswertungen nicht mehr genutzt werden.
13. Verfahren nach einem der Ansprüche 1 bis 12, dadurch gekennzeichnet, daß die Zuführung der Daten, die Anfrage und/oder die Ausgabe der Ergebnisse über das Internet erfolgt.
14. Computersystem zur Datenauswertung, gekennzeichnet durch:
 Mittel zur Eingabe von Erfahrungsdaten und objekt- oder personenspezifische Anfragedaten,
 Mittel, durch welche diese Anfragedaten unter Benutzung der Erfahrungsdaten bei einer auf die Anfragedaten bezogene Anfrage automatisch ausgewertet werden,
 Mittel zur automatischen Bewertung der Ergebnisse dieser Auswertung,
 Mittel, durch welche anhand dieser Bewertung, der zugehörigen Anfrage und den bei der Auswertung benutzten Daten automatisch ein Maß für den Einfluß dieser benutzten Daten auf die Auswertung ermittelt wird, und
 Mittel zur Datenausgabe.
15. Computerprogrammprodukt zur Datenauswertung, das direkt in den internen Speicher eines digitalen Computers geladen werden kann und Softwarecodeabschnitte umfaßt, mit denen das Verfahren gemäß Anspruch 1 ausgeführt wird, wenn das Produkt auf einem Computer läuft.
16. Computerprogrammprodukt zur Datenauswertung, das auf einem computergeeigneten Medium gespeichert ist und folgendes umfaßt:
 computerlesbare Programmmittel, die ein Computersy-

stem veranlassen, durch eine auf in dem Computersystem gespeicherte Anfragedaten bezogene Anfrage diese Anfragedaten unter Benutzung von in dem Computersystem gespeicherten Erfahrungsdaten automatisch auszuwerten,

computerlesbare Programmmittel, die ein Computersystem veranlassen, die Ergebnisse der Auswertung automatisch zu bewerten,

computerlesbare Programmmittel, die ein Computersystem veranlassen, anhand dieser Bewertung, der zugehörigen Anfrage und den bei der Auswertung benutzten Daten automatisch ein Maß für den Einfluß dieser benutzten Daten auf die Auswertung zu ermitteln.

17. Methode zur Nutzung eines Systems zur Auswertung von objekt- oder personenspezifischen Daten, dadurch gekennzeichnet, daß der Anbieter dieses Systems die Eingabe von und/oder den Zugang zu objekt- oder personenspezifischen Datensätzen für Nutzer des Systems mit einer gebührenpflichtigen PIN koppelt derart, daß die PIN gleichzeitig mit einem Mittel zur Erfassung von in das System einzugebenden objekt- oder personenspezifischen Daten und/oder einem Mittel zur Erfassung von Material, welches zur Ermittlung dieser objekt- oder personenspezifischen Daten dient, verbunden ist, und der Nutzer die PIN zusammen mit diesem (diesen) Mittel(n) durch Zahlung einer Gebühr erwirbt.

18. Methode nach Anspruch 17, dadurch gekennzeichnet, daß das Mittel zur Erfassung von Material ein Träger-Chip für die bei einem – erforderlichenfalls reproduzierbaren – Labortest benötigten Proben, wie z. B. ein DNA-Mikroarray, ist.

19. Methode nach einem der Ansprüche 17 oder 18, dadurch gekennzeichnet, daß der Nutzer des Systems zur Datenauswertung die PIN vom Anbieter oder einem vom Anbieter Beauftragten erwirbt.

Hierzu 10 Seite(n) Zeichnungen

40

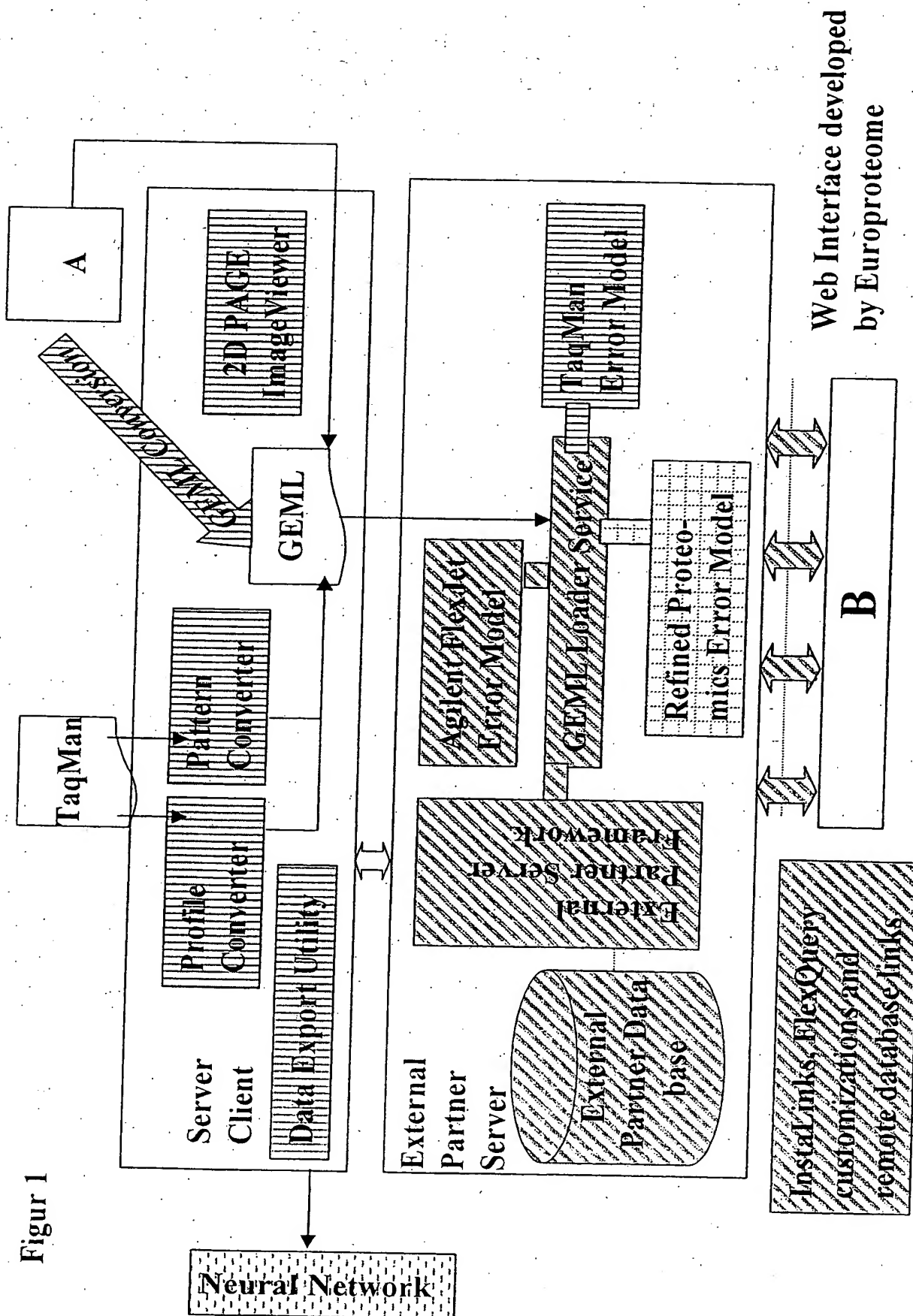
45

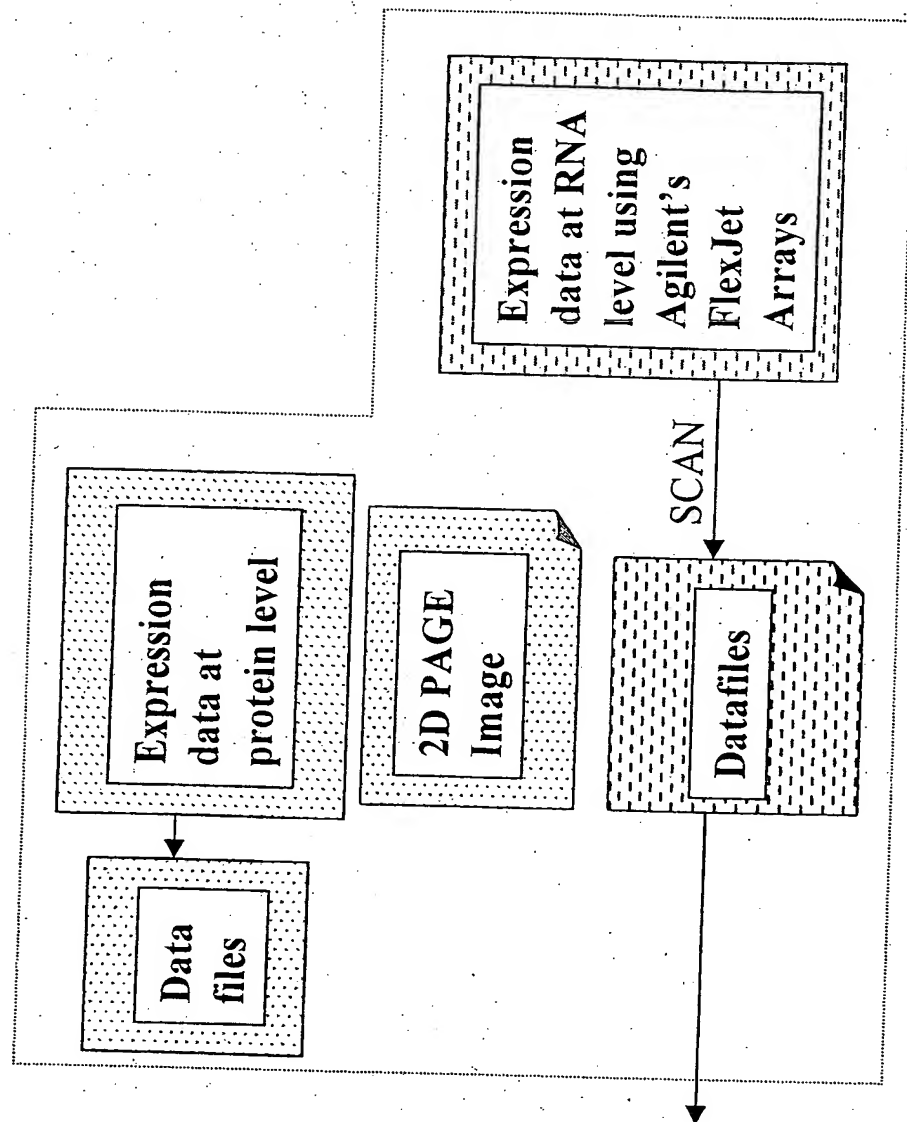
50

55

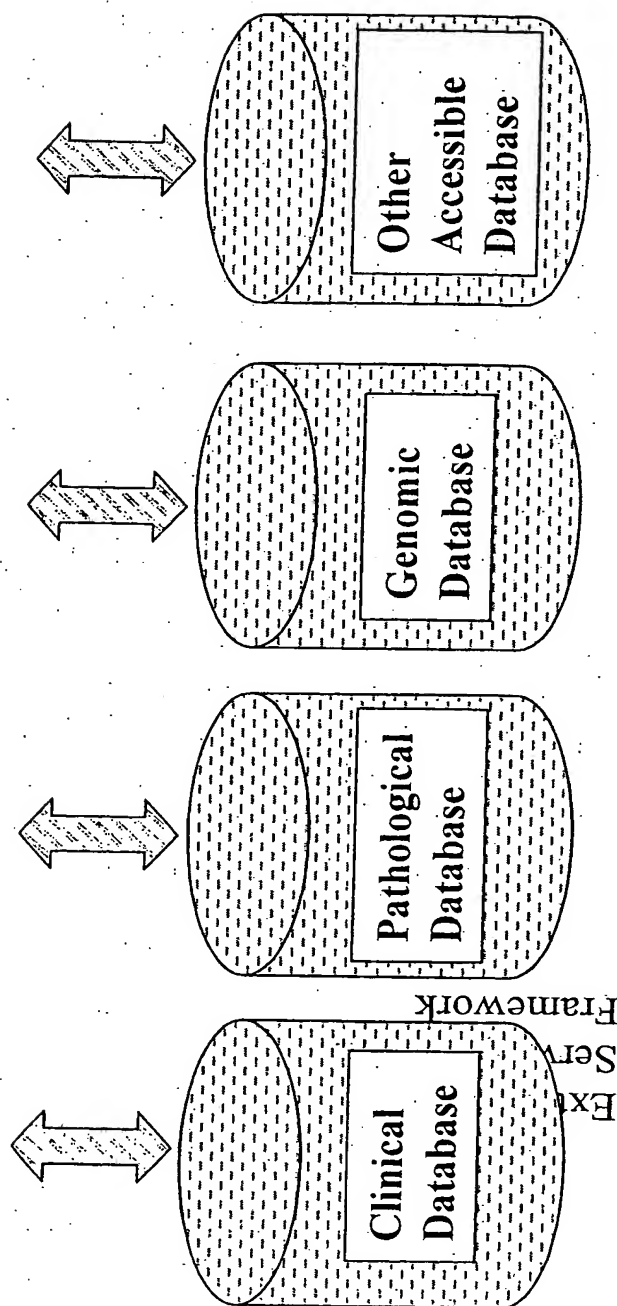
60

65

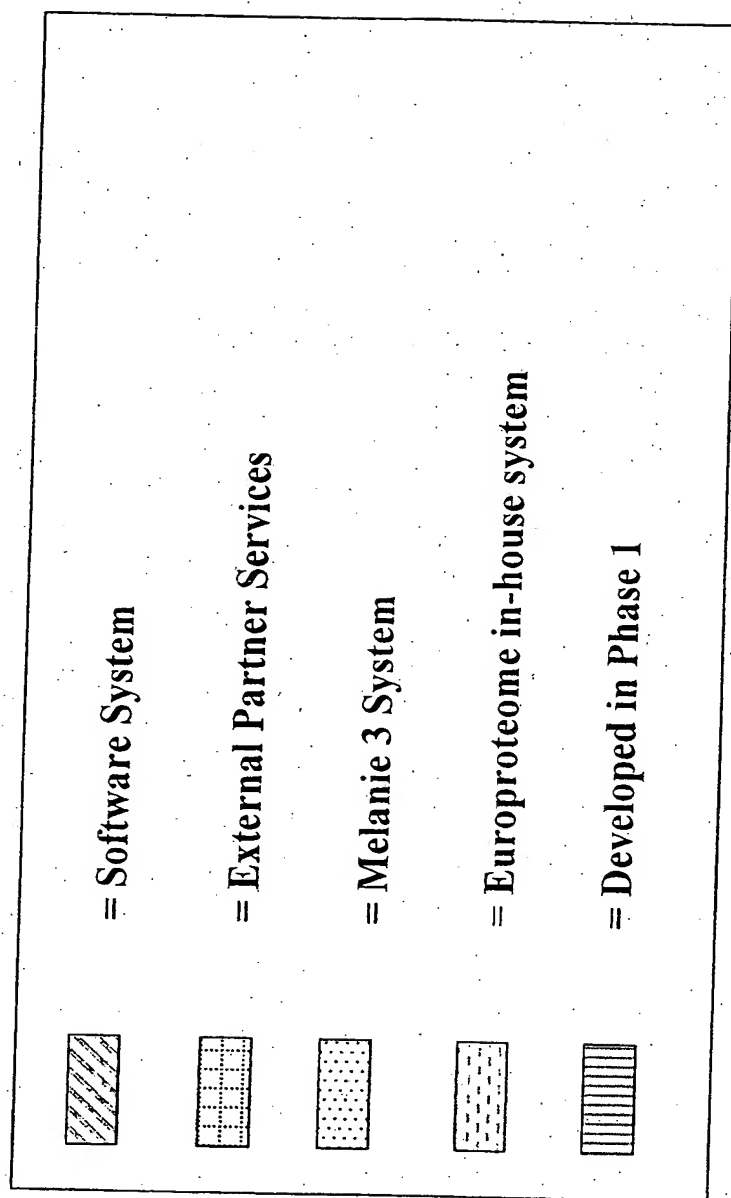




Teil A zu Figur 1



Teil B zu Figur 1



Legende zur Figur 1

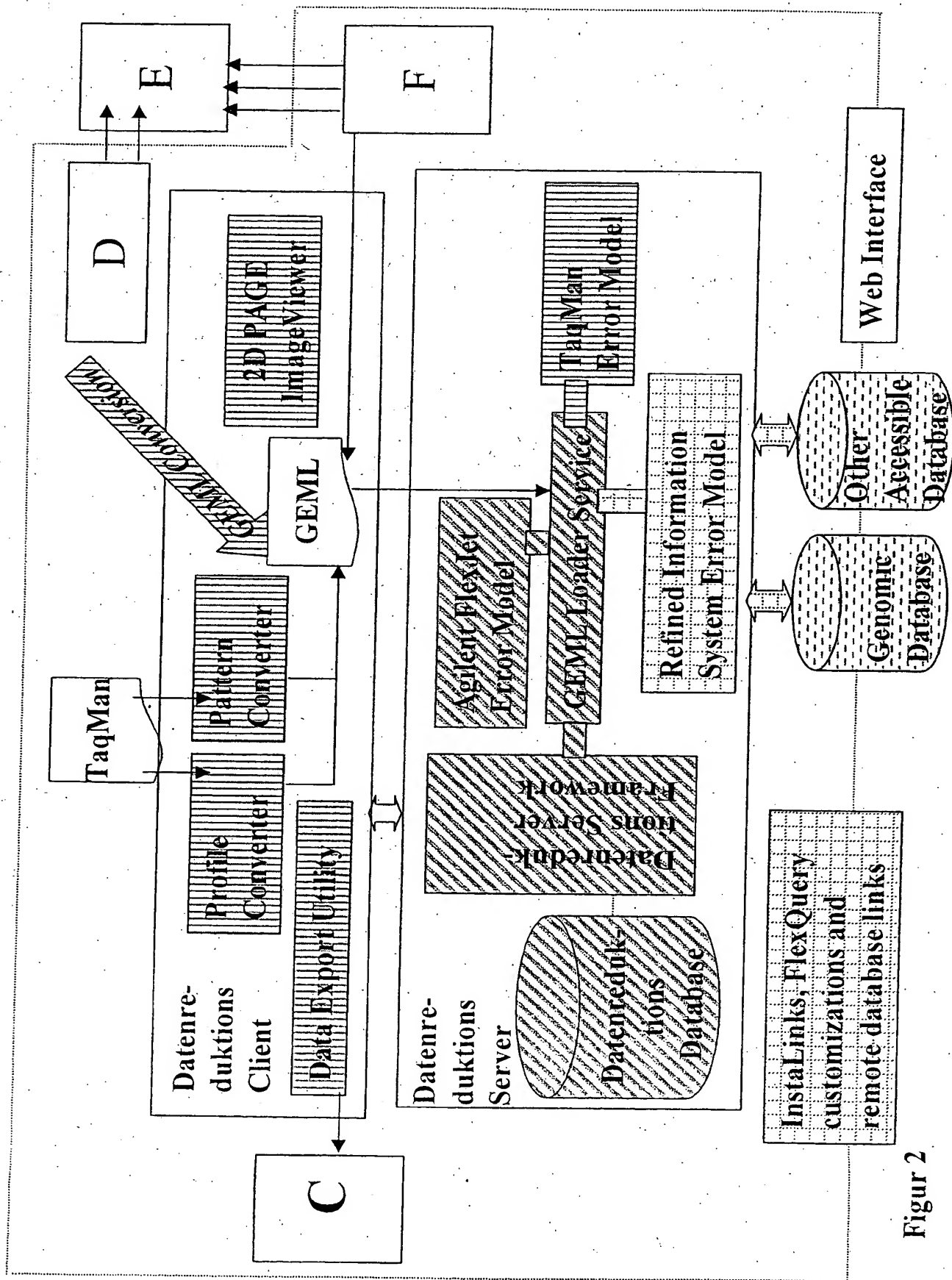
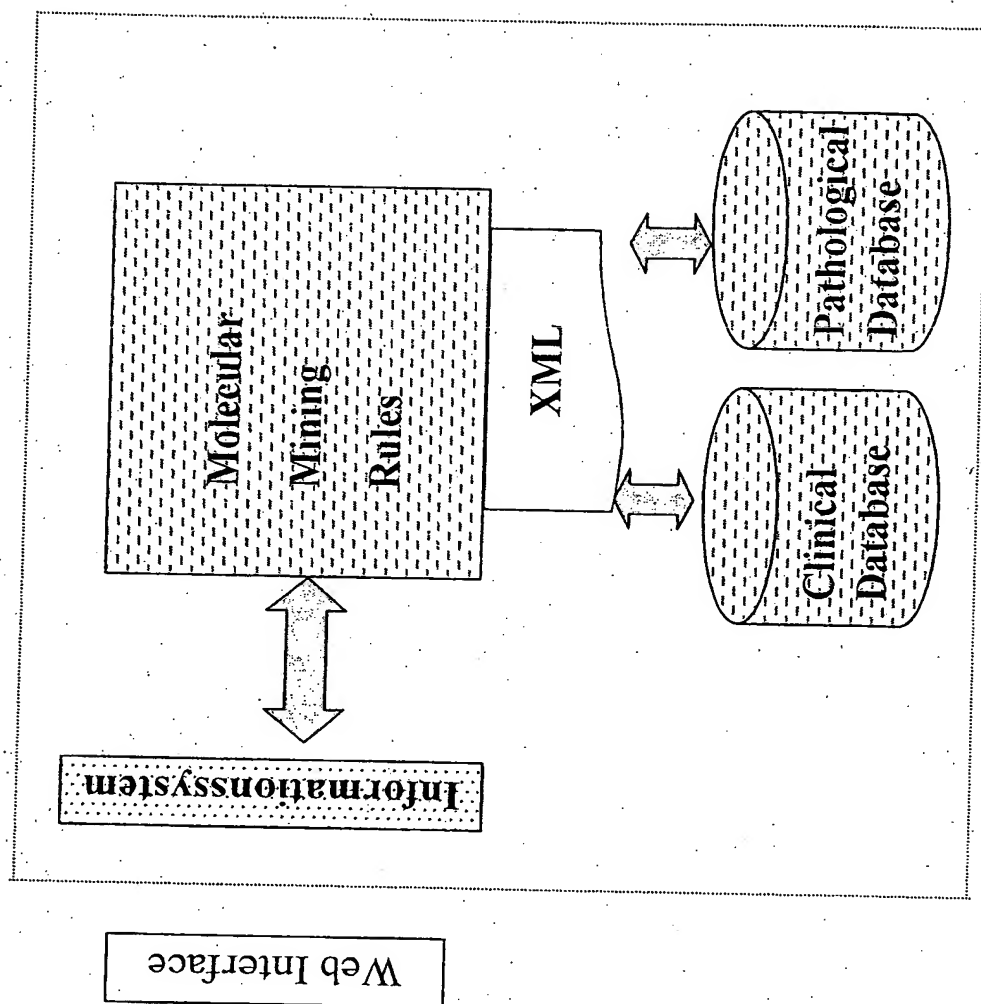
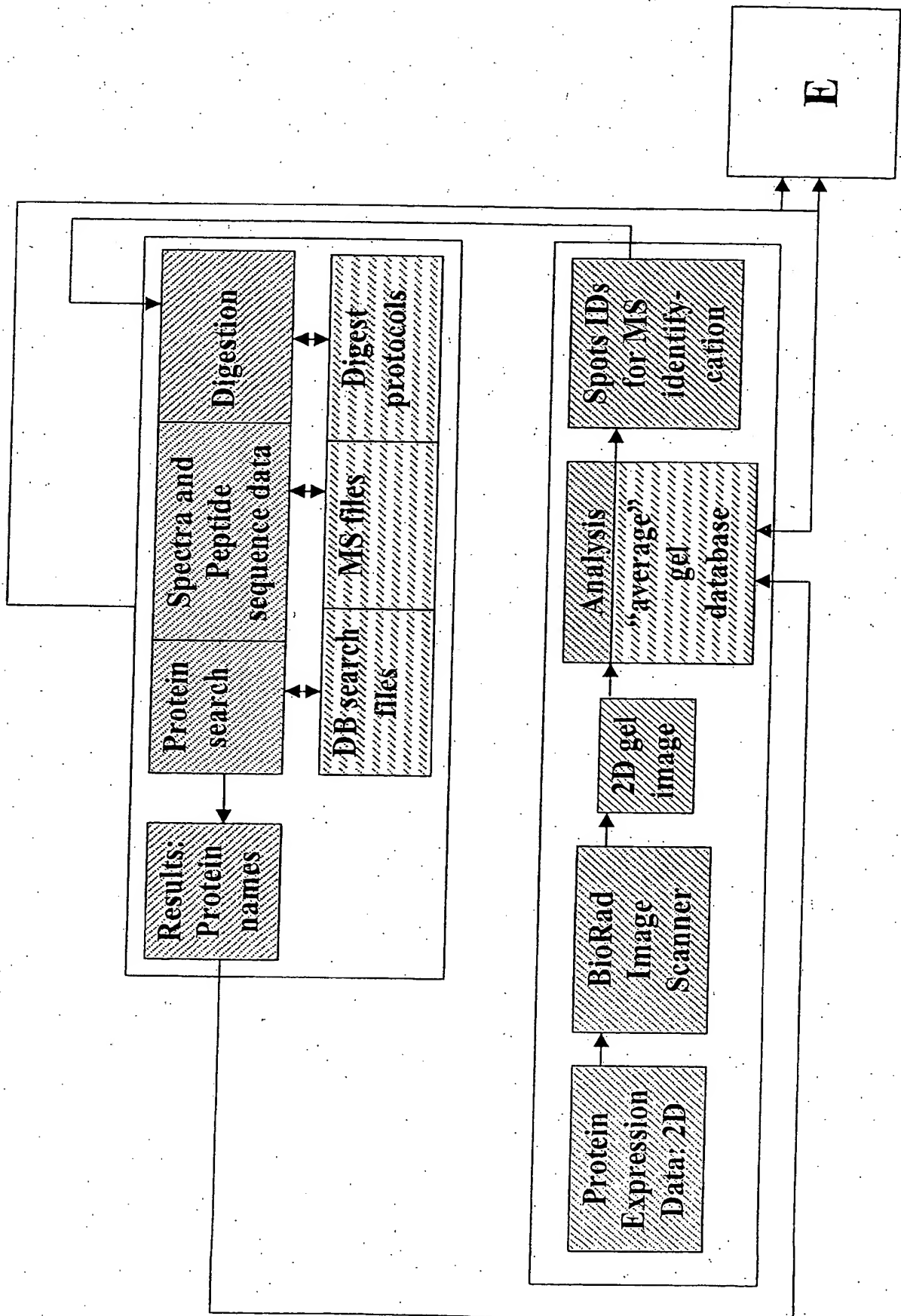


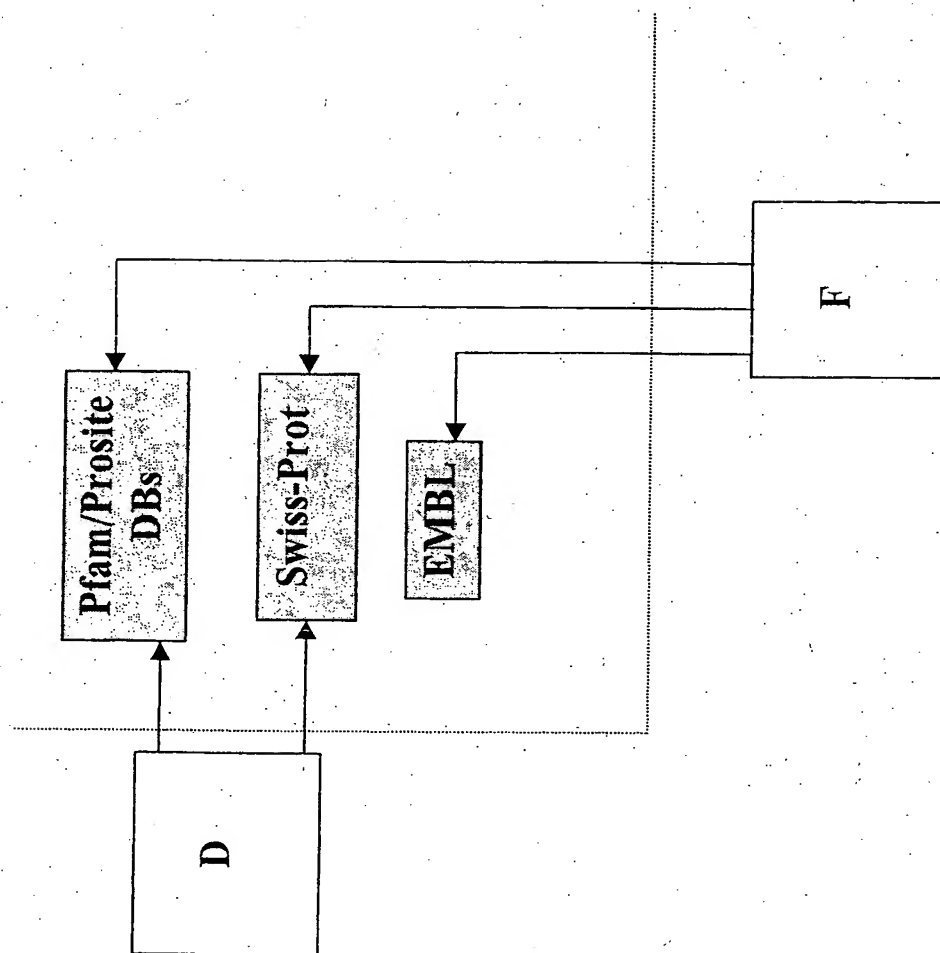
Figure 2



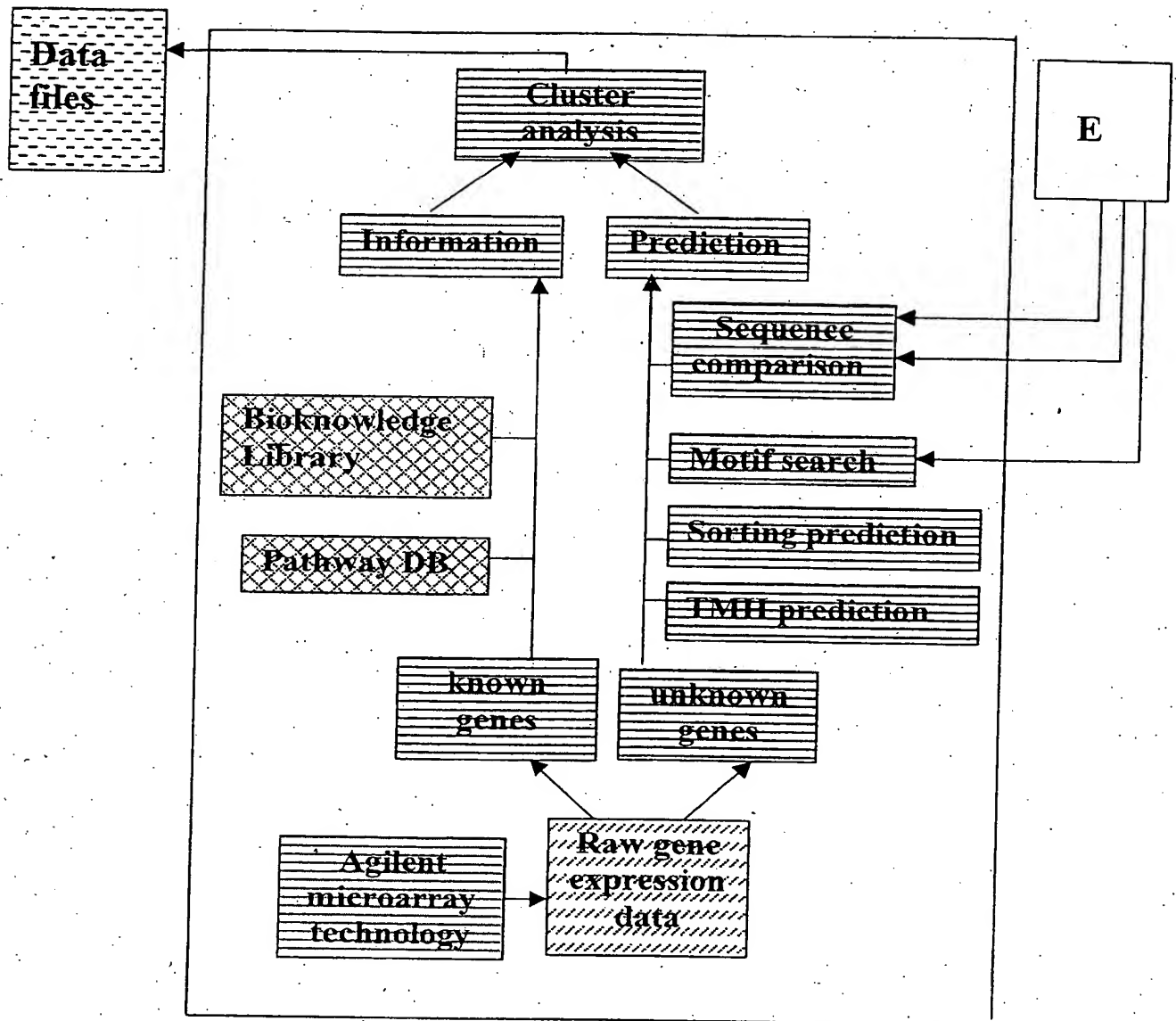
Teil C zur Figur 2



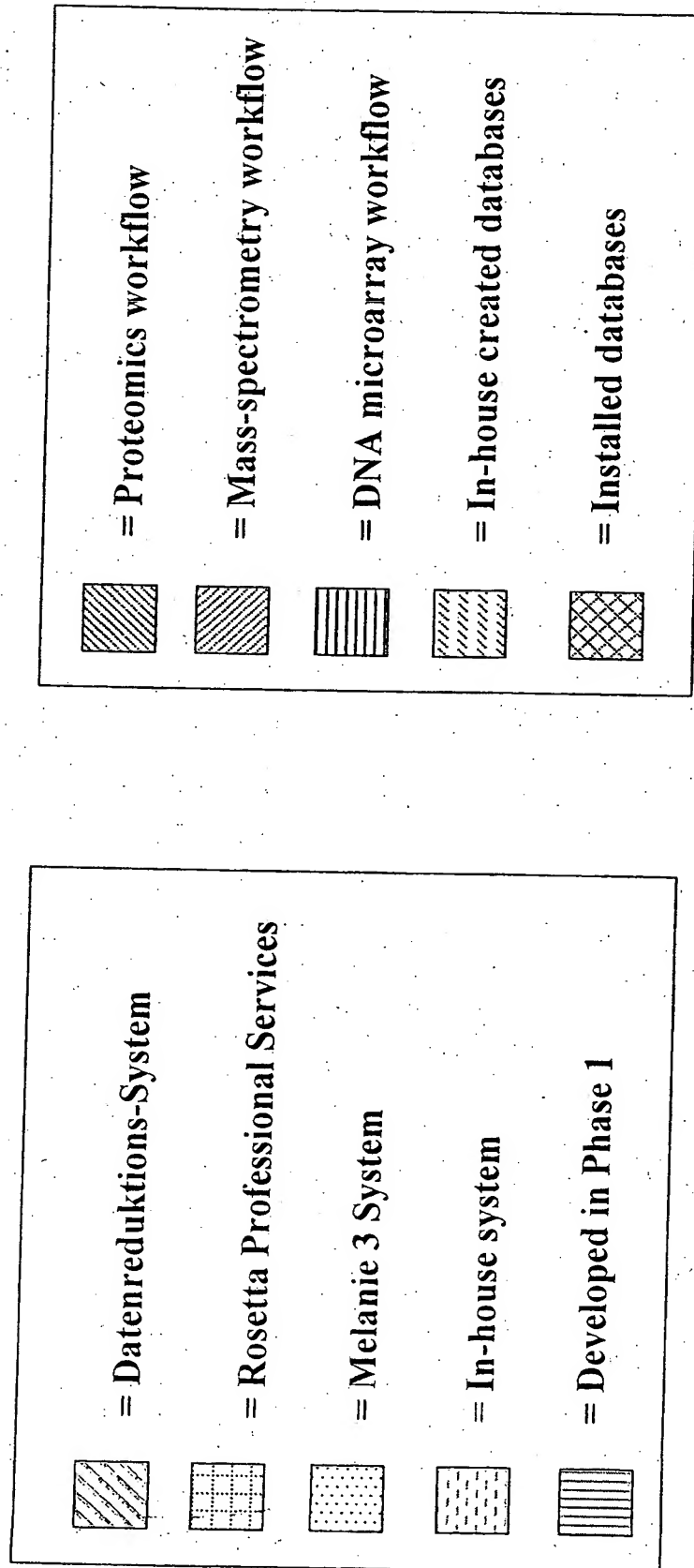
Teil D zur Figur 2



Teil E zur Figur 2



Teil F zur Figur 2



Legende zur Figur 2